Measuring continuous compositional change using decline and decay in zeta diversity

Running head: Turnover using zeta decline and decay

MELODIE A. MCGEOCH[1*], GUILLAUME LATOMBE[1], NIGEL R. ANDREW[2], SHINICHI NAKAGAWA[3], DAVID A. NIPPERESS[4], MARIONA ROIGÉ[5], EZEQUIEL M. MARZINELLI[6,7,8], ALEXANDRA H. CAMPBELL[6,7], ADRIANA VERGÉS[6,7], TORSTEN THOMAS[6], PETER D. STEINBERG[6, 7, 8], KATHERINE E. SELWOOD[9], MARIE V. HENRIKSEN[1] AND CANG HUI[10]

[1]*School of Biological Sciences, Monash University, Clayton, VIC 3800, Australia*

[2] *Zoology, University of New England, Armidale, NSW 2351, Australia*

[3]*Evolution and Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia, and Diabetes and Metabolism Division, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia*

[4]*Department of Biological Sciences, Macquarie University, NSW 2109, Australia*

[5]*National Centre for Advanced Bio-Protection Technologies, Lincoln University, Canterbury 7647, New Zealand. Current address: AgResearch, Private Bag 4749, Christchurch 8140, New Zealand.*

[6] *Centre for Marine Bio-Innovation, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia*

[7]*Sydney Institute of Marine Science, 19 Chowder Bay Rd, Mosman NSW 2088, Australia*

[8]*University of Sydney, School of Life and Environmental Sciences, Sydney, NSW 2006, Australia*

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ecy.2832

[9]*School of Biosciences, University of Melbourne, Parkville, VIC 3010, Australia and Wildlife and Conservation Science, Zoos Victoria, Parkville, 3052, Vic, Australia*

[10]*Centre for Invasion Biology, Department of Mathematical Sciences, Stellenbosch University, Matieland 7602, South Africa, and African Institute for Mathematical Sciences, Cape Town 7945, South Africa*

*Corresponding author: E-mail: melodie.mcgeoch@monash.edu

Abstract. Incidence, or compositional, matrices are generated for a broad range of research applications in biology. Zeta diversity provides a common currency and conceptual framework that links incidence-based metrics with multiple patterns of interest in biology, ecology and biodiversity science. It quantifies the variation in species (or OTU) composition of multiple assemblages (or cases) in space or time, to capture the contribution of the full suite of narrow, intermediate and wide-ranging species to biotic heterogeneity. Here we provide a conceptual framework for the application and interpretation of patterns of continuous change in compositional diversity using zeta diversity. This includes consideration of the survey design context, and the multiple ways in which zeta diversity decline and decay can be used to examine and test turnover in the identity of elements across space and time. We introduce the zeta ratio-based retention rate curve to quantify rates of compositional change. We illustrate these applications using 11 empirical datasets from a broad range of taxa, scales and levels of biological organisation – from DNA molecules and microbes to communities and interaction networks – including one of the original data sets used to express compositional change and distance decay in ecology. We show (i) how different sample selection schemes used during the calculation of compositional change are appropriate for different data types and questions, (ii) how higher orders of zeta may in some

cases better detect shifts and transitions, and (iii) the relative roles of rare versus common species in driving patterns of compositional change. By exploring the application of zeta diversity decline and decay, including the retention rate, across this broad range of contexts, we demonstrate its application for understanding continuous turnover in biological systems.

Keywords: DNA methylation, environmental gradients, diversity index, spatial and temporal turnover, species composition, interaction turnover, metagenome, microbial community, rare and common species.

INTRODUCTION

Changes in the patterns of species composition in space and time, along with richness, abundance and biomass, are critical to understanding what drives biodiversity and the ways that humans are transforming it (McGill et al. 2015, Socolar et al. 2016). Compositional change is relevant not only to species diversity, but to other levels of biological organisation, including molecular, genetic and clade diversity (e.g. Nipperess et al. 2012, Thomas et al. 2016), as well as to social phenomena (e.g. Vaz et al. 2017). The concept of turnover in the identity of elements in a system and its measurement is therefore of interest across a broad range of biological and socioecological contexts that span multiple scales (Arita et al. 2012, Shimadzu et al. 2015).

Compositional, or incidence-based, turnover is traditionally measured using metrics based on pairwise comparisons (i=2) of species incidence across sites or samples (Jost et al. 2010), commonly referred to as beta diversity (such as Jaccard, Sørensen or Simpson dissimilarity). Regularly used multisite metrics are based on combinations of multiple, pairwise measures (e.g. Diserud and Ødegaard 2007, Baselga 2013), and are also generally

applied such that they produce a single value of compositional turnover for $i=n$ sites. However, with information on only the alpha and all pairwise beta components in a community, it is not possible to express the full complement of biodiversity partitions across multiple sites (Hui and McGeoch 2014). In addition, differences in species composition between pairs of sites is mostly driven largely by rare species, whereas compositional similarity across larger number of sites tends to be determined by the more common species (which are, by definition, shared by large numbers of sites) (Latombe et al. 2017). As a result, comparisons of composition across greater than two sites ($i>2$) provides information on the contribution of increasingly more common species in the assemblage to compositional change.

Zeta diversity was introduced as a concept and metric that focusses attention on such multisite, cross-scale assemblage patterns of compositional change, with the purpose of better understanding how biodiversity change is structured (Hui and McGeoch 2014). Zeta diversity (ζ_i), which is the number of species shared by a given number of sites, provides a measure of turnover for each combination of i sites (Hui and McGeoch 2014). It draws attention to the information value of the full suite of multisite comparisons to quantifying compositional change, and enables the exploration of how incidence-based composition changes with the number of sites (i), scale, distance and time. Pairwise site calculations ($i=2$) of compositional similarity are encompassed within the zeta diversity framework. The applied value of zeta diversity has to date been shown in particular cases, for example as a measure of similarity and uncertainty in pest profile analysis (Roigé et al. 2017), to upscale estimates of biodiversity (Kunin et al. 2018), and to explore patterns of compositional change across streams (Simons et al. 2019), islands (Leihy et al. 2018) and protected areas (Britton et al. 2017).

There are three main points of difference between zeta diversity and existing measures of compositional change. 1. Raw values of zeta (not transformed or normalised) provide the full complement of partitions possible in an *n*-site assemblage. 2. The information value of zeta diversity derives from the form and rate of change in values of zeta across *n*-wise comparisons, and over time or distance. 3. It provides information on the relative contributions of rare to increasingly common species to turnover. The term turnover is used here in its broadest sense to mean change in the presence or absence of elements across sites or over time, regardless of the existence or not of environmental gradients and encompassing the richness difference, nestedness and replacement components (Legendre 2014) of compositional change. We therefore use 'turnover' as synonymous with compositional change, acknowledging it has a more narrow use in the BAS framework (Baselga 2010) to mean species replacement. The zeta diversity approach complements existing measures of compositional change in instances where there is interest in understanding patterns and processes of turnover in multi-wise assemblage comparisons. Combined with existing spatial regression techniques and environmental data measured at the *n*-sites, zeta diversity provides a means to differentiate drivers of compositional turnover for the full spectrum of species, from rare to common (Latombe et al. 2017). Such understanding is increasingly relevant to conservation strategy, with the need to manage the responses of both rare and common species to environmental change.

Here, using examples across a range of levels of biological organisation, we show how species across the continuous spectrum of occupancy contribute to biodiversity turnover, and how this can provide insight on the nature of biological heterogeneity. We start by (i) framing the calculation of turnover in the context of the structure of the data, and the selection of site combinations for calculation – all of which influence the outcome and interpretation of

patterns of turnover, and (ii) providing an overview of the calculation of zeta diversity and the range of patterns of compositional change it can be used to quantify. Building on zeta decline and zeta decay introduced in Hui and McGeoch (2014), we introduce the species retention rate, which quantifies relative rates of turnover across rare (low occurrence) to common (high occurrence) species. We then illustrate the suite of possible applications of zeta diversity to incidence matrices, and how zeta diversity should be applied and interpreted in each case.

CALCULATION OF ZETA DIVERSITY

Zeta diversity (ζ_i) is the number of species shared by *i* number of sites, with *i* referred to as the zeta order. It thus expresses compositional change in terms of similarity rather than dissimilarity. The average number of species across all sites (or alpha diversity) is ζ_1 (where $i = 1$), while the total observed or estimated richness across all sites or assemblages, as usual, represents gamma diversity. The units of measurement of zeta diversity (ζ_i) are therefore 'number of shared species', which has a minimum value of zero (no species shared by two or more sites) and a maximum value affected by the total species richness of the community matrix. Incidence-based, pairwise beta similarity metrics can all be derived from ζ_1 and ζ_2 (Hui and McGeoch 2014), and higher orders of zeta ($i>2$) represent the contribution of increasingly common species (widespread, with large area of occupancy) to compositional change. The number of orders (i.e. number of sites or cases) considered when calculating zeta is decided based on the dataset and question of interest, and at a maximum will be the total number of sites. If zeta reaches zero after *i* orders, i.e. no species is shared by more than *i* sites, there is of course no information to be gained by expressing it for orders beyond this.

For a community with a total number of *S* species discovered in *N* sites, zeta diversity can be calculated using the Monte Carlo method of resampling for a specific zeta order and sample selection scheme, while the mean and variance for the ALL sample selection scheme can also be estimated analytically. Specifically, let species *j* occur in n_j sites and let species *j* and *k* co-occur in $n_{j\cap k}$ sites. The probability of species *j* occurring in *i* sites is $P_j^i = C_{n_j}^i / C_N^i$ and the expected probability of species *j* and *k* co-occurring in *i* sites is $P_{jk}^i = C_{n_{j\cap k}}^i / C_N^i$, with C_n^i representing the binomial coefficient of *n* for chosen *i*; with the covariance $\mathrm{Cov}\left(P_j^i, P_k^i\right) = P_{jk}^i - P_j^i \times P_k^i$. The mean and variance of zeta diversity from the ALL sample selection scheme can then be calculated as $\mathrm{E}(\zeta_i) = \sum_{j=1}^{S} P_j^i$ and $\mathrm{Var}(\zeta_i) = \sum_{j=1}^{S}\sum_{k=1}^{S} \mathrm{Cov}(P_j^i, P_k^i)$.

The relationships between zeta diversity, sample properties, species accumulation curves and the species-area relationship derived using zeta diversity have been shown, representing its response to scale and sampling adequacy (Hui and McGeoch 2014). If desired, and bearing in mind the pros and cons of such metric transformation (Ulrich et al. 2018), raw zeta diversity can be normalised to express Jaccard, Sørensen and Simpson-equivalent indices, and partitioned into replacement, richness difference and nestedness indices (sensu Baselga 2010, 2013, Podani et al. 2013, Legendre 2014, Latombe et al. 2018). For example, let $\zeta_1(i)$ be the number of species in site *i*, $\zeta_2(i \cap j)$ the number of species shared by site *i* and *j*, $\zeta_3(i \cap j \cap k)$ the number of species shared by site *i*, *j* and *k*, and $S_3(i \cup j \cup k)$ the total number of species in site *i*, *j* and *k* ($= \zeta_1(i) + \zeta_1(j) + \zeta_1(k) - \zeta_2(i \cap j) - \zeta_2(i \cap k) - \zeta_2(k \cap j) + \zeta_3(i \cap j \cap k)$; Hui and McGeoch 2014). We thus have the normalised zeta of order 3 in the Sørenson-equivalent form, $\zeta_3(i \cap j \cap k)/\mathrm{mean}(\zeta_1(i), \zeta_1(j), \zeta_1(k))$; in the Simpson-equivalent form $\zeta_3(i \cap j \cap k)/\min(\zeta_1(i), \zeta_1(j), \zeta_1(k))$; and in the Jaccard-equivalent form $\zeta_3(i \cap j \cap k)/S_3(i \cup j \cup k)$ (Latombe et al. 2018). Normalised zeta of order

2 for these three forms is similar to the corresponding standard pairwise Sørenson, Simpson and Jaccard similarity indices.

Hypothesis testing and statistical inference with zeta diversity is similar to that used in ecology generally and in studies of compositional change using other metrics. Since ζ_i is the number of species shared by *i* sites, for each order of zeta the standard deviation can be computed from the Monte Carlo method of resampling (or based on the analytic formula of variance $\mathrm{Var}(\zeta_i)$ for the ALL sample selection scheme, see below) and plotted as a confidence interval to aid interpretation of the differences between the form and rate of turnover. Linear and nonlinear models using zeta values as the response variable and the order (zeta decline) or the physical distance (zeta decay) between sites as the predictors are fitted to the data (see below), and the 95% confidence intervals of the coefficients computed to assess their significance. Generalized linear and additive models, as well as spline regression, can be applied to test the response of species turnover (measured by zeta diversity) to environmental gradients (called multi-site generalized dissimilarity modelling; Latombe et al. 2017, 2018).

DATA STRUCTURE AND COMPOSITIONAL CHANGE

For any dataset the combination of a specific data structure and sample selection scheme results in alternative combinations for calculating turnover that require choices to be made prior to analysis, and consideration post analysis to interpret the patterns found (Appendix S1: Table S1). Study system dimensions and data structure determine not only the most appropriate choice of analysis but also how patterns in turnover are interpreted (Gotelli and Colwell 2001, Scheiner et al. 2011) (Fig. 1). Data structure constrains the range of biologically relevant ways in which the data may be selected (subsampled) during the process

of calculating turnover, i.e. the ways in which pairs, triplets and so on of sites are combined. The sample selection scheme may encompass all, a random selection or only a subset of possible combinations of *i* samples (Fig. 1). The main sample selection schemes are all combinations (ALL), nearest neighbours (NN), and fixed-point origin (FPO) or fixed-edge origin (FEO) (Fig. 1). Finally, as with the application of any diversity metric in ecology, the adequacy of the survey design and species richness estimates are key considerations when interpreting the outcome.

PATTERNS OF CONTINUOUS COMPOSITIONAL CHANGE

The two main applications for quantifying patterns in continuous compositional change are (1) zeta decline, including the species retention rate based on the zeta ratio, and (2) zeta decay over space or time (Table 1).

Zeta decline: continuous turnover with increasing numbers of sites

Zeta decline provides information on the form and rate of decrease in the average number of species shared across increasing number of sites (Table 1). It quantifies how the number of shared species decreases with zeta order, i.e. with increasing number of sites (cases or time periods) included in the calculation of shared species. Zeta decline is represented as a plot of zeta diversity values against the order of zeta, where orders represent selected pairs (order 2 for value zeta 2), triplets (order 3 for value zeta 3) of sites and so on (Hui and McGeoch 2014). When zeta is calculated using the ALL sample selection scheme (Fig. 1a, e), it provides an average expectation of compositional change in the data and could be considered as the upper bound (least shared species) of expected turnover against other structured sample selection schemes (e.g. NN, FPO and FEO). In cases where sites or surveys are positioned across a spatial or temporal gradient, and zeta is calculated using the nearest neighbours

sample selection scheme (NN for non-directional or directional options, Fig. 1 b,c,f,g,i), zeta diversity will typically decline at a comparatively slower rate. This is due to the constraints imposed by this spatial or temporal dependence in the sample selection scheme (versus the ALL scheme that considers combinations of sites that are randomly selected, and therefore less likely to share species than close sites). Other sample selection schemes may be envisaged for more specific applications.

Features of interest in zeta decline include: (i) the rate of decline in shared species, particularly across the first few orders, and (ii) if at higher orders the curve reaches or approximates zero, or not (Table 1). The larger the change in the value of zeta across subsequent orders, the greater the relative difference in the numbers of rare versus increasingly common species in the community. At lower orders this provides information on the rate at which relatively rare species no longer contribute to shared community membership as more sites are included. At higher orders, the value of zeta diversity provides information on the core species in the community, which is of interest itself but also for comparisons across datasets.

The parametric form of zeta decline, for example as best fit by either a power law or exponential relationship, provides information on the relative probability of species occurrences across sites, and may be used to test hypotheses about the degree to which biological matrices or communities are structured (Hui and McGeoch 2014). For instance, a null model where all species have an equal chance of occurring in any given site (although the sites themselves may be heterogeneous) will result in an exponential decline in zeta diversity. By contrast, a null model where each species has a unique probability of occurring at a site (generating a non-random pattern of co-occurrence), will lead to the power-law form

of zeta decline (Hui and McGeoch 2014). As such, estimated using ALL site combinations (Fig. 1a, e), the form of decline is interpreted as species having the same (exponential) or unequal (power law) probability of being observed across sites (see Kunin et al. 2018 for other possible forms). The goodness of fit of these forms can be compared using the Akaike Information Criterion.

Retention rate based on the zeta ratio

Knowing the rate at which rare and common species are retained across sites in a landscape or in surveys over time has intuitive ecological appeal as a measure of biodiversity change. Plotting the 'zeta ratios' (ζ_2/ζ_1, ζ_3/ζ_2, ζ_4/ζ_3 etc.) against the order of the zeta on the denominator (i.e. order 1 for the ratio ζ_2/ζ_1, order 2 for ζ_3/ζ_2 etc.) provides a retention rate curve that can be interpreted as (i) the rate (or the probability) at which species remain in the community as additional sites are included in the comparison, (ii) the chance of rediscovering (retaining) species in an additional site, and (iii) the degree to which common species are more likely to remain across sites than rare ones. Because common species are more likely to occur in extra samples than rare species (dependent to some extent on scale (grain) and species aggregation) (Harte 2008, Hui and McGeoch 2008), by comparing the ratios of zeta diversity values (e.g. ζ_{10}/ζ_9 vs. ζ_2/ζ_1), it is thus possible to assess the extent to which this is the case. Because the average number of shared species declines with increasing numbers of sites (as in zeta decline), a random species shared by *i* sites has a probability ζ_{i+1}/ζ_i of still being shared by *i*+1 sites (i.e. this species has a probability $1-\zeta_{i+1}/\zeta_i$ of being present in only *i* sites). For example, the zeta ratio for order nine is interpreted as the probability of retaining species that have at least an occupancy of nine (present at nine sites) in a tenth site, or the probability that these species become more widespread with the addition of another site. Examining zeta diversity expressed as a ratio is also useful for visualising turnover at higher orders (which in

zeta decline curves may be difficult to distinguish because the values of zeta are low) and highlights potential differences between the zeta declines of related datasets (Table 1).

Diversity indices scaled to lie between zero and one, such as the zeta ratio and retention rate curve, are useful for comparing studies or systems with different numbers of species and therefore potentially large differences in raw values. However, they are inevitably sensitive to species richness and show a loss of discriminant power when approaching the bounds (0, 1) (Gotelli and Colwell 2001, Baselga 2007). As such, this must be taken into account when interpreting normalised or transformed zeta metrics and we recommend that the retention rate is always interpreted alongside raw (untransformed) zeta decline.

Zeta decay: spatially or temporally explicit turnover

Zeta decay is conceptually similar to the distance decay of similarity (Nekola and McGill 2014), or species–time relationships and time decay (Shade et al. 2013), and provides information on the spatial or temporal extent of communities (i.e. the extent over which membership is shared). Zeta decay quantifies change in the number of species shared with increasing distance between sites (or time between surveys). Spatial and temporal compositional similarity for each order of zeta thus provides information on the form of decay for the rare to more common species in the community over time or distance. Zeta decay, or a plot of zeta diversity across sets of sites that are different distances or times apart, is represented with each zeta order as a different decay curve. Features of interest of zeta decay are (i) the shape and rate (slope) of decay, and how this differs across orders of zeta, (ii) the absolute distance (or time) over which this decay in shared species occurs, and (iii) the presence or absence of sharp declines in the curves (Table 1). Temporal decay curves

represent the change in number of shared species across subsequent surveys or time periods (this can vary with sample selection scheme, Fig. 1).

EXAMPLES

To demonstrate the application of zeta diversity and its interpretation, we selected eleven datasets representing a range of taxa, levels of biological organisation and spatial or temporal scales (Table 2). The concept of beta diversity was first demonstrated by R.H. Whittaker using tree community composition datasets (Whittaker 1960, 1967). We use one of these original datasets to illustrate the conceptual shift from beta to zeta diversity. In addition to six more standard species assemblage matrices, the examples used include one cellular, two molecular, one temporal and one interaction matrix. Ecological metrics are increasingly being used for other biological applications (e.g. Ma et al. 2019) and here we included a dataset on sub-cellular patterns of turnover, i.e. the incidence of DNA hypermethylation (a mechanism used by cells to control gene expression) at CpG nucleotide sites in tissues from patients with and without a metabolic disorder (Table 2). The question of interest here is whether the distribution of hypermethylation across nucleotide sites (i.e. compositional turnover) distinguishes patients with and without a developmental disorder.

Across the datasets, incidence therefore represents either the presence of a species or some other level of biological organisation or event (Table 2). We use the term species except when referring specifically to another incidence type (such as operational taxonomic units (OTUs)). In one dataset, incidence in the matrix represents the presence of an interaction rather than the presence of a particular species (i.e. to examine interaction turnover in a ‘gall wasp’ associated network). We also generally use the term site, which refers to locality, sample or assemblage, except where the data include cases (such as patients) or time periods

(for temporal turnover). We use 'community' or 'assemblage' to refer to the species by site matrix. Each data set was structured as a species by site matrix with non-zero marginal totals. Singletons were removed from some datasets, especially where they are likely to be a result of under sampling or sampling bias (Appendix S1: Table S2, Fig. S1). All analyses were done using the *zetadiv* package V.1.1 (Latombe et al. 2018), in R (R Core Team 2013) (Data S1).

1. Zeta decline results

Whittaker's (1956) tree data were used to show how traditional pairwise decline using Jaccard similarity compares with the decline in Jaccard-equivalent normalised zeta diversity for *n*-sites (Fig. 2). Normalised zeta across the transect was lower than pairwise Jaccard when using the fixed-point origin (FPO) sample selection scheme (Fig. 1) with the first elevational band as the fixed-point (the same is true for averaged Jaccard pairwise comparisons). Because zeta diversity also includes all sites located between the fixed-point and a specified elevational band in the calculation, the number of shared species will necessarily be lower than between the fixed-point and the specified elevational band (i.e. Jaccard). Moreover, using Jaccard-equivalent normalised zeta necessarily results in fewer shared species across subsequent orders (elevational bands), resulting in monotonic declines of zeta values. In contrast, the standard pairwise Jaccard similarity values can increase due to the reappearance of species in subsequent orders (elevational bands) (Fig. 2). This example thus illustrates the basis for differences between pairwise and multisite partitions of compositional change.

As outlined above, the choice of the combination of data structure and sample selection scheme will affect the outcome and is therefore important to consider *a priori* to ensure selection of the most appropriate combination for the data and hypothesis of interest (Fig. 1). To illustrate this effect we compare zeta diversity decline using three schemes on the

'Trees' and 'Sydney birds' data (Fig. 3a-b). The rate of decline over all combinations (ALL) and nearest neighbours (directional NN) is shallower compared to the fixed-point origin (FPO). This is a consequence of spatial clustering of species and the continuity of ranges, particularly of the more common species across the transect. This is apparent for both the one-dimensional 'Trees' data (Fig. 3a), and the two-dimensional data structure for 'Sydney birds' (Fig. 3b). Spatially or environmentally structured sample selection schemes thus affect the form of zeta decline, and may therefore be compared with the ALL scheme to test mechanistic explanations of turnover (McGill and Nekola 2010, Myers and LaManna 2016, Latombe et al. 2017).

In the following sections we examine the ecological interpretation of zeta decline (1.1), introduce species retention rate curves built from the zeta ratio (1.2), and discuss the form of zeta decline (1.3).

1.1 The ecological interpretation of zeta decline

To compare a diversity of datasets we used Sørensen-equivalent normalised zeta on 'Crop pest', 'DNAm', 'Bioregion birds' and the 'Soil metagenome' datasets (Table 2, Fig. 4). From Fig. 4a, it is apparent that in some cases the average number of species shared across sites declines to approximately zero within the extent of the study system, with very few widespread species shared by many sites, whereas in datasets with multiple species present in all sites, zeta converges towards this number of widespread species at high orders. Zeta diversity declined sharply for 'Crop pests', with almost complete turnover in the pest assemblage expected across more than 6 states or countries. Therefore, although there is a small suite of widespread insect crop pests globally shared by several countries, the global composition of pest assemblages actually differs widely (Roigé et al. 2017). This could either

suggest, for example, that most insect pests are crop specialists with narrow physiological niches, or that the current global invasion of crop pests has not yet reached an equilibrium. Although the rate of decline in the 'Soil metagenome' data at a micro scale was somewhat slower, it also declined to approximately zero after ~ 10 orders (Fig. 4a) (for microbial diversity studies, the pre-analysis data handling must be considered when interpreting diversity patterns, such as taxonomic resolution, sequence depth and treatment of singletons). Ecologically, the rapid zeta decline towards zero in both the crop pest and soil metagenome case, demonstrates strongly localised shared membership, albeit at very different scales.

By contrast, values of zeta > 0 at the highest expressed order (order 12 in Fig. 4) represent the average number of species (or % with normalised zeta) shared by large numbers of sites, i.e. the most common subset of species in the assemblage for that order. For example, zeta declines to approximately 20 % of bird species shared by Australian bioregions by order 12 (zeta of ~ 0.2 in Fig. 4a). There is therefore a core set of common bird species (~ 20% or 50 species, across 12 orders) shared across bioregions in this acse.

Intermediate to the other datasets in Fig. 4a, shared nucleotide sites at which hypermethylation occurs ('DNAm' data) has < 10% of OTUs shared by order 12. The zeta decline is steeper across higher orders (> 5) for the DNAm data than in the other examples, indicating rapid loss of common gene expression patterns among patients. The low percentage of shared nucleotide sites (low zeta diversity) at order 12 implies personalised gene expression via DNA methylation across these sites. It would also be affected in part by the pre-analysis removal of nucleotide sites at which hypermethylation occurred in all patients, because they are uninformative in the context of this dataset (Appendix S1: Table S2).

1.2 Retention rate based on the zeta ratio

The information value of retention rate curves as complementary to zeta decline is evident from the comparison of the results from three sample selection schemes on the 'Trees' dataset (Fig. 3c). The zeta retention rate curve is particularly striking for the DNN and FPO schemes, with a rapid decline in the rate of species retention beyond zeta orders 6-9 (Fig. 1c,d). This is not apparent from the decline in Jaccard-equivalent zeta similarity (Figs 2, 3a), nor from the zeta ratio using the ALL scheme (Fig. 3c). Although Whitaker (1967) appropriately interpreted the pattern of distance decay (Jaccard-based in Fig. 2) as the existence of 'broadly overlapping' tree species distributions across the transect, he also anecdotally pointed out the existence of a switch in dominance from cove forest species to gray beech and a suite of small tree species at ~ 1400 m a.s.l. along the transect (Whittaker 1960). This switch coincides with the abrupt shift in species composition between orders 6-8 (976 m -1098 m), detected by the zeta ratio and shown by the sharp decline in species retention rate for the directional and fixed-point origin (FPO) sample selection schemes (Fig. 3c). In the 'Trees' data, the retention rate of zeta diversity computed with the appropriate sample selection scheme thus enabled the identification of the ecotone noted by Whittaker (1960), by capturing the contribution of common species to turnover along the gradient, beyond the information provided by pairwise beta diversity (ζ_2). Using a different example, bird composition across Sydney shows no sudden shift for any of the three sample selection schemes (Fig. 3d). The rate of species retention stabilises beyond zeta order 10, demonstrating the absence of any conspicuous ecotone or dispersal barrier to bird species across the city to urban area of Sydney.

With a suite of examples for comparison (Fig. 4b), all species retention rates start by increasing, indicating a rapid loss of rare species between sites. This also shows that pairwise beta turnover is largely driven by the gain or loss of rare species (consistent with strong modes of rare species in the occupancy frequency distribution of these datasets, Appendix S1: Fig. S2). The probability of retaining common species is much lower for 'Crop pests' than for 'Bird bioregions', but the rates of common species retention for both these datasets start to asymptote beyond order 6 (Fig. 4b), suggesting a lower but equal chance of retaining common species (occurring in more than 6 sites) in these two cases. The retention rate for the 'DNAm' data increases and then starts to decline (i.e. showing signs of becoming modal - for a stronger example of this form of species retention curve see Fig. 5a-b). This means that at higher orders (> 7) there is a decline in the probability of retaining more common DNA methylation patterns among individual patients (i.e. a decrease in the rate of species retention) (Fig. 4b). The retention rate curves of DNAm and the soil metagenome intersect at ~ order 7, which is the order at which the retention rate of OTUs in the soil metagenome dataset becomes higher than that of the retention rate of shared nucleotide sites in the DNAm data.

1.3 The form of zeta decline

The 'DNAm' data were better fit by an exponential than power law (AIC -39.77 versus -18.93), whereas the difference was marginal for 'Crop pests' (AIC -1.96 exponential versus -1.47 power law) (Fig. 4c,d). This result shows, at least for the 'DNAm' data, a lack of structure in the matrix and that there are approximately equal probabilities of hypermethylation occurring at any nucleotide site. The 'Bioregion birds' and 'Soil metagenome' data were better fit by a power law (AIC value differences > 30) (Fig. 4c,d), demonstrating the existence of structure in these systems, and uneven probabilities in the occurrence of species (OTUs) across sites.

1.4 Within-system comparisons of zeta decline

In the previous examples (Fig. 4) we contrasted datasets that would not normally be included in the same study for the purpose of illustrating a range of possible forms of zeta diversity decline and retention rate. Zeta diversity is more likely to be applied to compare compositional change between taxa, functional groups, habitats or conditions within the same system. Here, using raw rather than normalised zeta, we explore three examples to compare zeta diversity between groups within the same system.

Example 1a: Does turnover differ between taxonomic groups of insect herbivores on a common host? Clear differences are apparent in compositional change between the two groups of '*Acacia* herbivores' on their host plant (Fig. 5a) (95% CI = [1.81, 1.94], linear model). The decline in beetles shared across sites is very rapid (and exponential, ΔAIC = 20.01), reaching a zeta diversity of zero by order 10, in contrast to slower decline in shared bugs across the same latitudinal range (power law, ΔAIC = 26.49) (Fig. 5a). Whereas the species retention rate in bugs is increasing, for beetles the retention rate drops beyond zeta order 5 (Fig. 5a). The probability of retaining beetle species in the assemblage (zeta ratio) beyond order 4-5 declines rapidly, suggesting complete turnover in the composition of beetles on *Acacia* within the extent of this study. Low prevalence and abundance of beetles in samples (Andrew and Hughes 2005) is a plausible explanation for the strong decline in species retention and lack of structure (i.e. exponential zeta decline) observed in these data for this taxonomic group in comparison with bugs.

Example 1b: Interaction turnover in a gall wasp-natural enemy network. The number of gall wasp-parasitoid interactions shared across sites in this network is higher and declines more slowly across zeta orders compared to interactions between gall wasps and inquilines

(Fig. 5b). No inquiline interactions are shared beyond the tenth order, whereas an average of ~ 3 parasitoid interactions remain shared at the same order. This difference is supported by the exponential fit of zeta decline for inquiline interactions (ΔAIC = 16.86), with little difference between a power and exponential fit for parasitoid interactions (ΔAIC = 1.34) (Fig. 5b). For inquilines the rate at which interactions were retained increases and then drops sharply from the third order, whereas it increases and drops gradually from order six for interactions involving parasitoids. These results are consistent with the expected ecological interpretation of inquilines as generalist herbivores opportunistically inhabiting the fleshy galls (and a strongly stochastic basis for the frequency with which they interact with different gall wasp species), compared with the host-specific relationships between gall wasps and parasitoids (Henriksen et al. 2017, 2018).

Example 1c: Does the distribution of hypermethylation across nucleotide sites (i.e. compositional turnover) distinguish patients with and without a developmental disorder? There was little difference in compositional turnover of the incidence of hypermethylation at nucleotide sites across patients with (form not distinguishable, ΔAIC= 0.01) and without (exponential, ΔAIC = 3.99) a metabolic disorder evident from a comparison of their zeta decline and retention rate curves (Fig. 5c). A linear model was computed for each order of zeta, including differences in age and disease status (0/1) as predictors and the zeta values computed for each combination of sites as the response variables (Latombe et al. 2017). Status was not significant (supporting the multivariate analysis-based findings of Ginsberg et al. 2012). However, age was a significant predictor of ζ_3 (95% CI = [-66.07, -25.79]) and ζ_4 (95% CI = [-58.19, -25.81]), but not ζ_2 (95% CI = [-98.44, 8.47]). Therefore, while the general prevalence of a relationship between DNA methylation and age is well known

(Horvath 2013), its effect on turnover was detected here only for orders of zeta greater than 2, i.e. not detected for pairwise comparisons (ζ_2).

2. Zeta decay results

Zeta decay, or a plot of zeta diversity across sets of sites that are different distances or times apart, is shown and interpreted for three examples below (Fig. 6), with the decline and retention rate results shown for comparison on the right of Fig. 6.

Example 2a: Spatially explicit turnover in alien and native plants. The comparative distances (km) in zeta values across orders 3-5, especially at larger distances in Fig. 6b, show that there are more ubiquitous species (both locally and regionally widespread) in the alien than the native community (Fig. 6a,b). The overall patterns of distance decay in this dataset thus reinforce the interpretation of zeta decline in Fig. 6c, i.e. more shallow compositional turnover in aliens than natives (in Fig. 6c, clear differences are apparent in the zeta declines of (95% CI = [1.74, 1.95], linear model). Here, however, the difference in rates of decline are calibrated against distance, enabling scale-specific comparisons of distance decay across species groups. Over distances of 20 km, on average there are 2 - 6 alien species shared (across zeta orders), whereas there is around one native species shared by sites (Fig. 6a,b). Shared alien species decline more slowly than shared native species for ζ_2, whereas for orders ζ_3 the slope of the zeta decays are similar for aliens and natives despite the difference in absolute zeta values. The retention rate curve confirms that in the alien community, common species are more likely to be retained across sites (by between ~40-70%) than in the native community (~10-40%) (Fig. 6d). Although there are over half as many alien as native plants in this region (Banks Peninsula, New Zealand, Wilson 2009), there are higher values of zeta

diversity (more aliens in common than natives) and slower turnover in alien compared to native plants, especially for rare species.

Example 2b: Kelp microbe communities across the eastern and western bioregions of Australia. Zeta decay curves encompassing one or more clear shifts (where the width of the error intervals should be broadly bounded by the amplitude of the shift or period) suggest the presence of a dispersal barrier, a shift in environmental conditions, patchiness or temporal periodicity of some form (Nekola and White 1999). For example, the steep decline in average numbers of shared kelp microbes (both rare to more common, i.e. from zeta order 2 upwards) over distances of 150-300 km along the coast of NSW suggests marked patchiness in assemblage structure at this scale (Fig. 6e). By contrast, the rate of distance decay in WA is shallow and consistent across the different orders, in spite of high total and average OTU richness in the region (Fig. 6g). On average, the number of shared species is higher and more consistent with distance in WA (total richness 550 species, mean ± s.d. = 346.88±23.49) compared with NSW (518 species, 288.33±60.02). The striking difference between decay curves (Fig. 6e-f) could be explained by distinctly different current systems between the coasts that drive the dispersion of kelp microbes in different ways (Thompson et al. 2011), although differences in the relative distances across surveyed sites may also play a role in this case (Appendix S1: Fig. S3).

Example 2c: Temporal decay of bird communities in catchments variously affected by a regional drought. Temporal decay was very different between the two catchments (Fig. 6i-j), and the zeta decline and retention rate results (right hand panels in Fig. 6) provide some insight on the differences observed (left hand panels in Fig. 6). Zeta decline at Catchment 2 is faster across low orders than at Catchment 1 (Fig. 6k), even though the mean richness

between them is similar (14.81±3.14 at Catchment 1; 12.18±3.31 at Catchment 2). This explains the absolute differences in zeta diversity values between Fig. 6 i-j, i.e. high at Catchment 1 and lower at Catchment 2. Shared species declined slightly over the full period at Catchment 2 and very little over short time periods, whereas it declined comparatively rapidly over short periods at Catchment 1 (< 250 days - left of decay curve in Fig. 6i). This was true of rare and more common components of the community (i.e. for the zeta orders 2 to 5 shown). There were relatively fewer species shared by the end of the period at Catchment 1 than at the start, whereas at Catchment 2 the number of shared species was more similar at the start and end of the period (i.e. turnover was comparatively independent of time, suggesting greater stability in the community under average rainfall conditions). The difference in number of surveys between catchments had little effect on this result (Appendix S1: Fig. S4). This comparison shows that turnover was fast and consistent under average rainfall, and lower and more variable over the period under severe drought.

DISCUSSION

Here we have provided a framework for placing studies of compositional change into their appropriate design context, and from this the range of questions about biotic heterogeneity that can be addressed. Together, zeta decline and zeta decay provide the full spectrum of diversity partitions for any given number of sites over a specified space or time period. The diverse range of empirical examples used show that zeta decline, the zeta ratio and retention rate, and zeta decay provide a range of insights on the nature of continuous compositional turnover, biodiversity structure or, in molecular and cellular studies, the form of biological heterogeneity. Quantifying this full spectrum of rare to intermediate and common species as they contribute to driving compositional change was more informative than pairwise turnover alone, particularly for comparisons of groups or communities within a study system.

The aim of a study will determine whether or not factors such as species richness gradients, species abundance or density, variation in the size of the species pool, interactions and abiotic environmental variables are either biodiversity drivers of primary interest, or confounding factors that must be accounted for to ensure unbiased inference (Gotelli and Colwell 2001, Andrew et al. 2012), as well as the most appropriate survey design to use to generate the data to begin with (Scheiner et al. 2011). The framework of system contexts (dimensions and sample selection approaches) used demonstrates the importance of biologically-driven decisions on how to treat the data pre-analysis, as well as the study specificity of how zeta decline is interpreted. For example, in datasets where a large proportion of the species are shared by the majority of sites (and where the value of zeta would therefore be high at high orders), it may be appropriate to consider this subset of species with a close to saturated distribution as uninformative and to exclude them – as we did for the high proportion of nucleotide sites at which hypermethylation occurred across all patients. These species (nucleotide sites in this case) may otherwise hide the signal in compositional change from the whole suite of less common species.

By contrast, in some systems focus on the common suites of species may itself be of interest (McGeoch and Latombe 2016, Baker et al. 2019). For example, in microbial studies the dynamics of 'core microbiomes' is meaningful (Shade and Handelsman 2012), and wide-ranging components of assemblages are also relevant in invasion biology (Leihy et al. 2018). Here the difference in zeta decay found between native (steep) and alien (shallow) plants is in the direction that one might expect given the tendency for alien plant species to have broader niches (Lockwood et al. 2005). Therefore, if on-ground surveys were to extend beyond the sample extent of this study, one might expect to discover new native (more rare) species at a faster rate than new alien species (with the assumption that local species richness remains

similar in the newly surveyed sites). Common and dominant species are important drivers of both diversity patterns and ecosystem function (Baker et al. 2019) and their roles are often contrasted with those of rare species (e.g. Draper et al. 2019). The concepts of rarity and commonness are however relative and lie on a continuum (McGeoch & Latombe 2017). The use of zeta diversity avoids the need for subjective thresholds being imposed to separate rare from common components of communities, and enables more informed analysis of the contribution of species to turnover across the full spectrum from rare to intermediate to common.

We introduced the use of the zeta ratio to express and compare retention rate curves that demonstrates the rate at which common species remain in, or are lost across, sites in a community. By highlighting turnover amongst species shared by multiple sites, we showed how retention curves can provide insight on shifts in dominant species, habitat use differences between taxonomic groups and interactions involving different functional groups. Three broad types of retention rate curves were distinguished (Figs 4b, 5 and 6). The biological significance of these will be study-dependent, but can generally be interpreted as follows. (i) An increasing curve shows that common species are more likely to be retained in additional samples than rare ones, and as a result perhaps that the sampling extent is smaller than the metacommunity, or that site selection is relatively homogenous and well characterised by habitat specialists (Myers and LaManna 2016). A clear example was provided by the retention rate curve for bugs in the *Acacia* herbivores data. (ii) In an asymptotic curve, an asymptote of 1 indicates the presence of common species over all sites, whereas an asymptote < 1 indicates that common and intermediate species are equally likely to be retained in subsequent sites or samples, e.g. turnover across sites in parasitoid-gall wasp interactions. (iii) A modal curve shows that for high orders of zeta, the most common species

are less likely to be retained with the addition of sites, i.e. the study extent encompasses a distinct community, metacommunity or spans an ecological boundary. Modal retention rate curves were apparent in 'Trees', *Acacia* herbivore (beetle) and gall wasp-inquiline interactions. Within the full extent of a community or system the three types of retention rate are likely to be a continuum, shifting from increasing to asymptotic if a core set of common species remain for a particular zeta order, and shifting either directly from increasing to decreasing, or via a modal curve, when moving beyond the footprint of the most common suite of species in the community.

The full spectrum of compositional turnover as captured by zeta diversity of different orders captures the patterns of both species occurrence (occupancy) and co-occurrence in the community (in the ALL sample selection scheme). The form of zeta decline can thus be indicative of ecological processes as reflected by the multispecies co-occurrence patterns that it quantifies. While forms other than exponential or power law are possible and may also be informative, as the two most regularly observed forms (Hui & McGeoch 2014), we examined the exponential and power law forms. Comparatively equal probabilities of the occurrence of species across sites (exponential form) have been suggested to be associated with stochastic assembly processes, whereas niche differentiation processes are more likely to produce a power law form of zeta decline in natural communities (for comparable mechanistic beta-diversity based interpretations see Munoz et al. 2008, Nekola and McGill 2014, Leihy et al. 2018). The fit can also be used to test the scale dependence of species incidence in the community (related to the hierarchical scaling of zeta diversity; Hui and McGeoch 2014); exponential reflects scale independence of species retention across sites, whereas the power law reflects non-independence of species composition across sites, and an increasing probability of retaining more common species at finer scales (Hui and McGeoch 2008,

McGlinn and Hurlbert 2012). As with any inference of process from pattern in ecology, clear hypotheses and strong inference approaches should be used to support the interpretation of the form of zeta decline in this way.

Here we largely used the diversity of examples and data structures to illustrate the application and interpretation of zeta diversity (rather than to test data set-specific hypotheses). They illustrate the rich information content of multi-site partitioning over pairwise contrasts, including the detection of a switch in dominance, differences in the turnover of interactions between specialists and generalists, and differences in features of the distribution of alien and native plant communities. Moving beyond quantification of patterns in turnover and the inferences that may be drawn from these, there is substantial interest in drivers of compositional change (Chase et al. 2018, D'Amen et al. 2018, Hui and Richardson 2019). The drivers of differences in bird community turnover shown between catchments, for example, clearly warrants further investigation. High productivity vegetation such as that found on floodplains has been found to buffer bird assemblages from the effects of drought in this region and may play a role in explaining site-by-site differences in species turnover during drought (Selwood et al. 2015, 2018). With additional catchment replicates this could be tested using the zeta diversity framework.

In spite of substantial focus on biodiversity change over the recent period, trends in spatial and temporal turnover across scales, from local to global, remain poorly supported by conceptual frameworks and empirical studies (Dornelas et al. 2013, McGill et al. 2015). Being able to disentangle spatial or temporal trends in rare to common species has significant potential value, given the importance of species composition to the delivery of ecosystem services (Baker et al. 2018). Here we showed how zeta diversity can contribute to filling this

gap when used to study trends in turnover across multiple sites. Future progress in modelling and hypothesis testing of continuous compositional change will be made using combinations of empirical and simulation modelling. The broad range of applications and insights that can be derived using zeta diversity on any incidence matrix (along with the development of related abundance and weighted matrix approaches) will we hope contribute to further development of general theory on the scaling and structure of biological heterogeneity.

ACKNOWLEDGEMENTS

We thank David Warton, Mark Westoby, Steven Chown, David Baker, Sarah Reeve, David Clarke, Fangliang He and Pierre Legendre for comments on the work. M.M., G.L. and C.H. acknowledge support from the Australian Research Council's *Discovery Projects* funding (project number DP150103017), N.R.A. from the Australian Research Council's *Discovery Projects* funding (project number DP160101561), and C.H. from the National Research Foundation of South Africa (no. 89967). S.N. is supported by a Future Fellowship (FT130100268). M.R. received financial support from the European Commission (Erasmus Mundus partnership NESSIE, ref. 372353-1-2012-1-FR-ERA MUNDUS-EMA22) and BPRC travel grant 2014. Refer to Appendix S1: Table S2 for comprehensive data use acknowledgements.

LITERATURE CITED

Andrew M. E., M. A. Wulder, N. C. Coops, and G. Baillargeon. 2012. Beta-diversity gradients of butterflies along productivity axes. Global Ecology and Biogeography 21: 352-364.

Andrew, N. R., and L. Hughes. 2005. Arthropod community structure along a latitudinal gradient: Implications for future impacts of climate change. Austral Ecology 30:281-297.

Arita, H. T., A. Christen, P. Rodriguez, and J. Soberon. 2012. The presence-absence matrix reloaded: the use and interpretation of range-diversity plots. Global Ecology and Biogeography 21:282-292.

Baker D. J., S. T. Garnett, J. O'Connor, G. Ehmke, R. H. Clarke, J. C. Z. Woinarski, and M. A. McGeoch. 2019. Conserving the abundance of nonthreatened species. Conservation Biology 33: 319-328.

Barrett, G., A. Silcocks, S. Barry, R. Cunningham, and R. Poulter. 2003. The New Atlas of Australian Birds. Birds Australia, Royal Australasian Ornithologists Union, Melbourne, Australia.

Baselga, A. 2007. Disentangling distance decay of similarity from richness gradients: response to Soininen et al. 2007. Ecography 30:838-841.

Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. Global Ecology and Biogeography 19:134-143.

Baselga, A. 2013. Multiple site dissimilarity quantifies compositional heterogeneity among several sites, while average pairwise dissimilarity may be misleading. Ecography 36:124-128.

Britton A. W., J. J. Day, C. J. Doble, B. P. Ngatunga, K. M. Kemp, C. Carbone, and D. J. Murrell. 2017. Terrestrial-focused protected areas are effective for conservation of freshwater fish diversity in Lake Tanganyika. Biological Conservation, 212:120-129.

Chase, J. M., B. J. McGill, D. J. McGlinn, F. May, S. A. Blowes, X. Xiao, T. M. Knight, O. Purschke, and N. J. Gotelli. 2018. Embracing scale-dependence to achieve a deeper understanding of biodiversity and its change across communities. Ecology Letters 21:1737-1751.

D'Amen, M., H. K. Mod, N. J. Gotelli, and A. Guisan. 2018. Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence. Ecography 41:1233-1244.

Diserud, O. H., and F. Odegaard. 2007. A multiple-site similarity measure. Biology Letters 3:20-22.

Dornelas, M., A. E. Magurran, S. T. Buckland, et al. 2013. Quantifying temporal change in biodiversity: challenges and opportunities. Proceedings of the Royal Society B-Biological Sciences 280:20121931.

Draper, F. C., G. P. Asner, E. N. H. Coronado, et al. 2019. Dominant tree species drive beta diversity patterns in western Amazonia. Ecology 100:e02636.

Ebach, M. C., A. C. Gill, A. Kwan, S. T. Ahyong, D. J. Murphy, and G. Cassis. 2013. Towards an Australian bioregionalisation atlas: A provisional area taxonomy of Australia's biogeographical regions. Zootaxa 3619:315-342.

Gotelli, N. J., and R. K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4:379-391.

Ginsberg, M.R., Rubin, R.A., Falcone, T., Ting, A.H. and Natowicz, M.R. 2012. Brain transcriptional and epigenetic associations with autism. PLoS ONE 7:e44736.

Harte, J. 2008. From spatial pattern in the distribution and abundance of species to a unified theory of ecology: The role of maximum entropy methods. Mathematical Modelling of Biosystems, pp. 243-272, Springer Berlin Heidelberg.

Henriksen M. V., D. G. Chapple, S. L. Chown, and M. A. McGeoch. 2017. Gall wasp biocontrol of invasive *Acacia longifolia*: implications of strong bottom-up effects. Ecosphere 8:e02043.

Henriksen M. V., D. G. Chapple, S. L. Chown, M.A. McGeoch. 2019. The effect of network size and sampling completeness in depauperate networks. Journal of Animal Ecology 88:211-222.

Horvath, S. 2013. DNA methylation age of human tissues and cell types. Genome Biology 14:R115.

Hui, C., and M. A. McGeoch. 2008. Does the self-similar species distribution model lead to unrealistic predictions? Ecology 89:2946-2952.

Hui, C., and M. A. McGeoch. 2014. Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns. American Naturalist 184:684-694.

Hui, C., and D. M. Richardson. 2019. How to invade an ecological network. Trends in Ecology and Evolution 34:121-131.

Hurst, J. M., and R. B. Allen. 2007. A permanent plot method for monitoring indigenous forests – Field protocols. Landcare Research, Lincoln, New Zealand.

Jost, L., A. Chao, and R. L. Chazdon. 2010. Compositional similarity and β beta. diversity. Biological Diversity Frontiers in Measurement and Assessment, pp. 66-84. Oxford University Press, Oxford.

Kunin, W. E., J. Harte, F. L. He, C. Hui, R. T. Jobe, A. Ostling, C. Polce, A. Sizling, A. B. Smith, K. Smith, S. M. Smart, D. Storch, E. Tjorve, K. I. Ugland, W. Ulrich, and V. Varma. 2018. Upscaling biodiversity: estimating the species-area relationship from small samples. Ecological Monographs 88:170-187.

Latombe, G., C. Hui, and M. A. McGeoch. 2017. Multi-site generalised dissimilarity modelling: using zeta diversity to differentiate drivers of turnover in rare and widespread species. Methods in Ecology and Evolution 8:431-442.

Latombe G., M. A. McGeoch, D. Nipperess, and C. Hui. 2018. zetadiv: an R package for computing compositional change across multiple sites, assemblages or cases. bioRxiv. doi:10.1101/324897

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology 74:1659-1673.

Legendre P. 2014. Interpreting the replacement and richness difference components of beta diversity. Global Ecology and Biogeography 23:1324-1334.

Leihy, R. I., G. A. Duffy, and S. L. Chown. 2018. Species richness and turnover among indigenous and introduced plants and insects of the Southern Ocean Islands. Ecosphere 9:e02358.

Lockwood, J. L., P. Cassey, and T. Blackburn. 2005. The role of propagule pressure in explaining species invasions. Trends in Ecology and Evolution 20:223-228.

Ma, Z., L. Li, and N. J. Gotelli. 2019. Diversity-disease relationships and shared species analyses for human microbiome-associated diseases. The ISME Journal. doi:10.1038/s41396-019-0395-y.

Marzinelli, E. M., A. H. Campbell, E. Z. Valdes, et al. 2015. Continental-scale variation in seaweed host-associated bacterial communities is a function of host condition, not geography. Environmental Microbiology 17:4078-4088.

Michael, C. A., M. R. Gillings, A. J. Holmes, L. Hughes, N. R. Andrew, M. P. Holley, and H. W. Stokes. 2004. Mobile gene cassettes: A fundamental resource for bacterial evolution. American Naturalist 164:1-12.

McGeoch, M. A., and G. Latombe. 2016. Characterizing common and range expanding species. Journal of Biogeography 43:217-228.

McGill, B. J., and J.C. Nekola. 2010. Mechanisms in macroecology: AWOL or purloined letter? Towards a pragmatic view of mechanism. Oikos 119:591-603.

McGill, B. J., M. Dornelas, N. J. Gotelli, and A. E. Magurran. 2015. Fifteen forms of biodiversity trend in the Anthropocene. Trends in Ecology and Evolution 30:104-113.

McGlinn, D. J., and Hurlbert, A. H. 2012. Scale dependence in species turnover reflects variance in species occupancy. Ecology 93:294-302.

Munoz, F., P. Couteron, and B. R. Ramesh. 2008. Beta diversity in spatially implicit neutral models: A new way to assess species migration. American Naturalist 172:116-127.

Myers, J. A., and J. A. LaManna. 2016. The promise and pitfalls of beta-diversity in ecology and conservation. Journal of Vegetation Science 27:1081-1083.

Nekola, J. C., and B. J. McGill. 2014. Scale dependency in the functional form of the distance decay relationship. Ecography 37:309-320.

Nekola, J. C., and P. S. White. 1999. The distance decay of similarity in biogeography and ecology. Journal of Biogeography 26:867-878.

Nipperess, D. A., A. J. Beattie, D. P. Faith, S. G. Ginn, R. L. Kitching, C. A. M. Reid, T. Russell, and L. Hughes. 2012. Plant phylogeny as a surrogate for turnover in beetle assemblages. Biodiversity and Conservation 21:323-342.

Podani, J., C. Ricotta, and D. Schmera. 2013. A general framework for analyzing beta diversity, nestedness and related community-level phenomena based on abundance data. Ecological Complexity 15:52–61.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. URL http://www.R-project.org/.

Roigé, M., M. A. McGeoch, C. Hui, and S. Worner. 2017. Cluster validity and uncertainty assessment for self-organizing map pest profile analysis. Methods in Ecology and Evolution 8:349-357.

Scheiner, S. M., A. Chiarucci, G. A. Fox, M. R. Helmus, D. J. McGlinn, and M. R. Willig. 2011. The underpinnings of the relationship of species richness with space and time. Ecological Monographs 81:195-213.

Selwood, K. E., J. R. Thomson, R.H. Clarke, M.A. McGeoch, and R. Mac Nally. 2015. Resistance and resilience of terrestrial birds in drying climates: do floodplains provide drought refugia? Global Ecology and Biogeography 24:838-848.

Selwood, K. E., M. A. McGeoch, R. H. Clarke, and R. Mac Nally. 2018. High-productivity vegetation is important for lessening bird declines during prolonged drought. Journal of Applied Ecology 55:641-650.

Shade, A., J. G. Caporaso, J. Handelsman, R. Knight, and N. Fierer. 2013. A meta-analysis of changes in bacterial and archaeal communities with time. The ISME Journal 7:1493-1506.

Shade, A., and Handelsman, J. 2012. Beyond the Venn diagram: the hunt for a core microbiome. Environmental Microbiology 14:4-12.

Shimadzu, H., M. Dornelas, and A. E. Magurran. 2015. Measuring temporal turnover in ecological communities. Methods in Ecology and Evolution 6:1384-1394.

Simons, A. L., R. Mazor, E. D. Stein, and S. Nuzhdin. 2019. Using alpha, beta, and zeta diversity in describing the health of stream- based benthic macroinvertebrate communities. Ecological Applications 29:e01896.

Socolar, J. B., J. J. Gilroy, W. E. Kunin, and D. P. Edwards. 2016. How should beta-diversity inform biodiversity conservation? Trends in Ecology and Evolution 31:67-80.

Thomas, T., L. Moitinho-Silva, M. Lurgi, et al. 2016. Diversity, structure and convergent evolution of the global sponge microbiome. Nature Communications 7:11870.

Thompson, P. A., P. Bonham, A. M. Waite, L. A. Clementson, N. Cherukuru, C. Hassler, and M. A. Doblin. 2011. Contrasting oceanographic conditions and phytoplankton

communities on the east and west coasts of Australia. Deep-Sea Research Part II-Topical Studies in Oceanography 58:645-663.

Ulrich, W., Y. Kubota, B. Kusumoto, A. Baselga, H. Tuomisto, and N. J. Gotelli. 2018. Species richness correlates of raw and standardized co-occurrence metrics. Global Ecology and Biogeography 27:395-399.

Vaz, A.S., C. Kueffer, C. A. Kull, et al. 2017. The progress of interdisciplinarity in invasion science. Ambio 46:428-446.

Whittaker, R. H. 1956. Vegetation of the Great Smoky Mountains. Ecological Monographs 26:1-69.

Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs 30:280-338.

Whittaker, R. H. 1967. Gradient analysis of vegetation. Biological Reviews 42:207-264.

Wilson, H. 2009. Natural History of Banks Peninsula, Canterbury University Press, Christchurch, New Zealand.

TABLE 1. Summary of the ways in which zeta diversity (ζ_i) is expressed and interpreted and the purposes for which it can be applied.

Expression	Interpretation	Features of interest	Application
Zeta diversity **decline**	**Analytical**: The change in number of species shared with increasing numbers of sites included in the comparison (the zeta order). **Ecological**: The contribution of narrow to wide-ranging species to compositional change (i) implicitly over space, time or cases (e.g. samples, sites or hosts), and (ii) zeta diversity is incidence-based and the form of rarity and commonness it captures is therefore the species range, occurrence or area of occupancy.	(i) The form of decline (e.g. exponential or power), (ii) the rate of decline, (iii) the order at which zeta reaches or approximates zero, and (iv) the value of zeta at the maximum sampled extent or the maximum order for which it is calculated.	(i) To test the degree to which a community is non-randomly structured, (ii) to test hypotheses about stochastic versus mechanistic determinants of community or system structure, and (iii) to compare or test hypotheses about the expected rate of compositional change, by comparison with a null expectation or between taxa, habitats or conditions of interest.
Zeta **ratio**	**Analytical**: The probability of retaining (or rediscovering) a species of the same order of commonness in additional cases. The order here refers to the zeta order or number of sites included in the zeta diversity calculation. **Example**: The probability of a species remaining present ('being retained') in nine sites with the addition of a 10^{th} site.	When expressed in the form of species retention rate curves (see below).	Used to construct the species retention rate.
Retention rate (based on the zeta ratio)	**Analytical**: The degree to which common species are more likely to remain ('be retained') across sites than rare ones with an increase in zeta order. **Ecological**: The rate at which species remain ('are retained') in the community across sites, or the rate at which common or widespread species are retained across the landscape.	The shape of the retention rate curve as either increasing, asymptotic, modal, or decreasing.	(i) For visualising turnover at high orders where absolute changes are small, (ii) to assess the spatial extent of a community or metacommunity relative to the sample extent, (iii) to compare or test differences in species retention rates between

			taxonomic groups, habitats or conditions of interest or against the null expectation, and (iv) to detect ecotones or abrupt shifts in composition.
Zeta diversity **decay** (related to distance decay)	**Analytical**: Change in the number of species shared with increasing distance between sites (or time between surveys) for different orders of zeta.	(i) Rate of decay, and (ii) differences in the slope of decay between zeta orders.	To select appropriate spatial and temporal dimensions when designing survey and monitoring schemes.
	Ecological – spatial: Distance decay in the compositional similarity of communities.	(i) Distance at which zeta approximates zero, and (ii) the value of zeta at particular distances of interest or at the maximum distance (study extent).	(i) To quantify the distances over which the composition of communities or systems change, and (ii) to compare the distances over which the composition of rare versus common components of communities or systems turn over.
	Ecological – temporal: Temporal change in compositional similarity of communities.	(i) Time period over which zeta approximates zero, and (ii) time periods of interest over which temporal change takes place.	(i) To quantify the time period over which the composition of communities or systems change, and (ii) to compare the time periods over which the rare and common components of communities or systems turn over.

TABLE 2. Properties of the 11 datasets used to illustrate the application of zeta diversity (in the form of species by site matrices, see Appendices S1, S3, S5-S6 for further detail). Species may be represented by other operational taxonomic units (OTUs) and sites by other types of sampling units.

Dataset	**Description**	**No. species (OTUs)**	**No. sites# [groups]**	**Grain**	**Spatial extent**	**Data structure, sample selection scheme (Fig. 1, a-j)**
1. 'Trees'[1]	Tree community composition along an elevational gradient across mesic sites in the Great Smoky Mountains surveyed by R.H. Whittaker.	39	11	120 m elevational bands	landscape – 1200m extent	One dimensional: ALL (a), NN (c) and FPO (d)
2. 'Sydney birds'[2]	A selection of atlas data for terrestrial (non-freshwater) species centred around Sydney, Australia).	145	22	25 x 25 km	regional – 150 km radius	Two dimensional: ALL (e), NN (g) and FPO (h)
3. 'Crop pests'[3]	Occurrence records at the level of country, state (province) and island group for insect pest species of interest to global crop protection.	868	373	'region' represented by a country or state	global	Two dimensional: ALL (e)
4. 'DNAm'[4] (with, without disorder) (human donor)	The incidence of DNA (hyper) methylation ("DNAm") at CpG nucleotide sites in human occipital cortex tissue from patients of a range of ages, with or without a developmental disorder (autism) (see Ginsberg et al. 2012). In this case, the 'species' are nucleotide sites and the tissue from individual patients are the cases (or 'sites').	1545	16 (8, 8) [2]	donor/host, tissue from human individuals	n/a	Non-dimensional: ALL combinations

5. 'Bioregion birds'[5]	Checklist-type lists of species across unequal area bioregions in Australia.	641	85	bioregions	biogeographic	Two dimensional: ALL (e)
6. 'Soil metagenome'[6]	Data from 5 ml soil samples from a spatial array in a dry sclerophyll woodland in Australia (Michael et al. 2004, see for further details on DNA extraction and gene cassette size class screening, assessment and characterisation). The data matrix used is based on small, mobile genetic elements (or gene 'cassettes') as OTUs and soil samples as 'sites'.	451	12	5 ml soil sample	Micro (local), 50 m^2	Two dimensional: ALL (e)
7. *'Acacia* herbivores' (beetles and bugs)[8]	Insect herbivores (beetles (Coleoptera) and bugs (Hemiptera)) sampled from a single host plant (*Acacia falcate*) at sites spanning a latitude range in Eastern Australia.	184 (74, 110)	12 [2]	groups of trees	biogeographic, 1200 km	Two dimensional: ALL (e)
8. 'Gall wasp interactions', (parasitoids and inquilines)[11]	Interactions between three gall wasp species (*Trichilogaster* spp.) and their natural enemies (parasitoids and inquilines) in Melbourne, Australia, involving a total of 34 interactions.	34 (15, 19)	13 [2]	circular sites, 1 km in diameter	regional, 525 km^2	Non-dimensional: ALL
9. 'Plants' (alien, native)[7]	Survey data from Banks Peninsula (New Zealand) of native and alien plant species from a regular array of plots approximately 1km apart across the extent of the Peninsula.	910 (316, 594)	1281 [2]	20 x 20 m plots	regional, ~50 x 30 km.	Two dimensional: ALL (e)
10. 'Kelp microbes' (NSW, WA)[9]	Microbial communities (bacterial and archaeal OTUs defined based on a <97% identity of their 16S rRNA genes) associated with the surfaces of	903 (518, 550)	17 [2]	kelp blades within regions and sites in each marine	biogeographic	Two dimensional: ALL (e)

	blades of common kelp (*Ecklonia radiata*) along the coastline of temperate Australia, including two marine Biogeographic Provinces (BPs) (alongside the Australian states of New South Wales (NSW) and Western Australia (WA)) (Marzinelli et al. 2015).			Biogeographic Province		
11. 'Temporal birds'[10]	Temporal data for native birds in two catchments in a major river basin in southeastern Australia, coinciding with a regional drought.	71, 56	6 (1998-2003) [2]	2 ha plots surveyed multiple times a year	local	One dimensional: FPO (d)

#Sites, hosts or temporally repeated surveys in the case of dataset 10. References describing the data set and/or the system context: 1. Table 5 in Whittaker (1956), 2. Barrett et al. (2003), 3. Roigé et al. (2017), 4. Horvath (2013), 5. Ebach et al. (2013), 6. Michael et al. (2004), 7. Hurst and Allen (2007), 8. Andrew and Hughes (2005), 9. Marzinelli et al. (2015), 10. Selwood et al. (2015), 11. Henriksen et al. (2017, 2018).

FIGURE LEGENDS

FIG. 1. One- and two- dimensional data structures (left) and alternative data sample selection schemes (right) for estimating compositional turnover using zeta diversity. Data may include broad geographic regions encompassing spatially homogenous or heterogeneous environments, independent units hosting a community (e.g. islands, hosts of parasite or bacterial communities or genomes) or linear habitats (e.g. coastlines or ecotones). Non-directional schemes (a-b, e-f) are those where no, or no single, environmental or spatial gradient is of concern or interest (sample units may also be discrete with their relative spatial position of no interest, i.e. non-dimensional). Directional structures (c-d, g-j) are those where there are known or designed directional gradients of interest, e.g. a one or two dimensional change in environmental condition away from a point source (d), gradients perpendicular to an edge or ecotone (f), or a time series or transect along an environmental gradient (i,j). 'ALL' represents all possible combinations of pairwise, triplet etc. site combinations. Cell or point shades represent different environmental conditions, e.g. in the directional schemes, the black cell or point represents the origin, fixed-point or -edge. The lines and arrows between sites (right) are not comprehensive and simply show how sites may be combined for the calculation of zeta diversity.

FIG. 2. Decline in compositional similarity of tree species along an elevation gradient in the Great Smoky Mountains, quantified using standard pair-wise Jaccard similarity (Whittaker 1967) and compared with normalised, n-wise Jaccard-equivalent zeta diversity decline (i.e. zeta diversity decline). Both elevational bands (equivalent to the zeta order in this case) and the distance along the elevational transect (m) can be shown on the x-axis in this specific case. Only in this particular and simple case of a one-dimensional data structure and a directional point source sample selection scheme, is zeta order (elevational bands in this case)

directly comparable to distance (change in elevation) along the transect ('Trees, Table 2), i.e. in this case zeta decline is directly comparable to zeta decay. The data underlying Fig. 2 match the scheme in Fig. 1d.

FIG. 3. Patterns of compositional change (zeta diversity decline (a, b) and the zeta ratio plotted as species retention rate(c, d)) for two directional data structures with different data sample selection schemes (ALL, DNN, FPO) (as shown in Fig. 1). Data sets used are trees along an elevation transect (a, c; 'Trees' Table 2) and bird communities radiating out from central Sydney (b, d; 'Sydney birds' Table 2). Data combination schemes: ALL, all combinations of n sites, directional nearest neighbour, FPO, fixed-point origin (see Fig. 1). The legend in panel (d) applies to all panels (a-d).

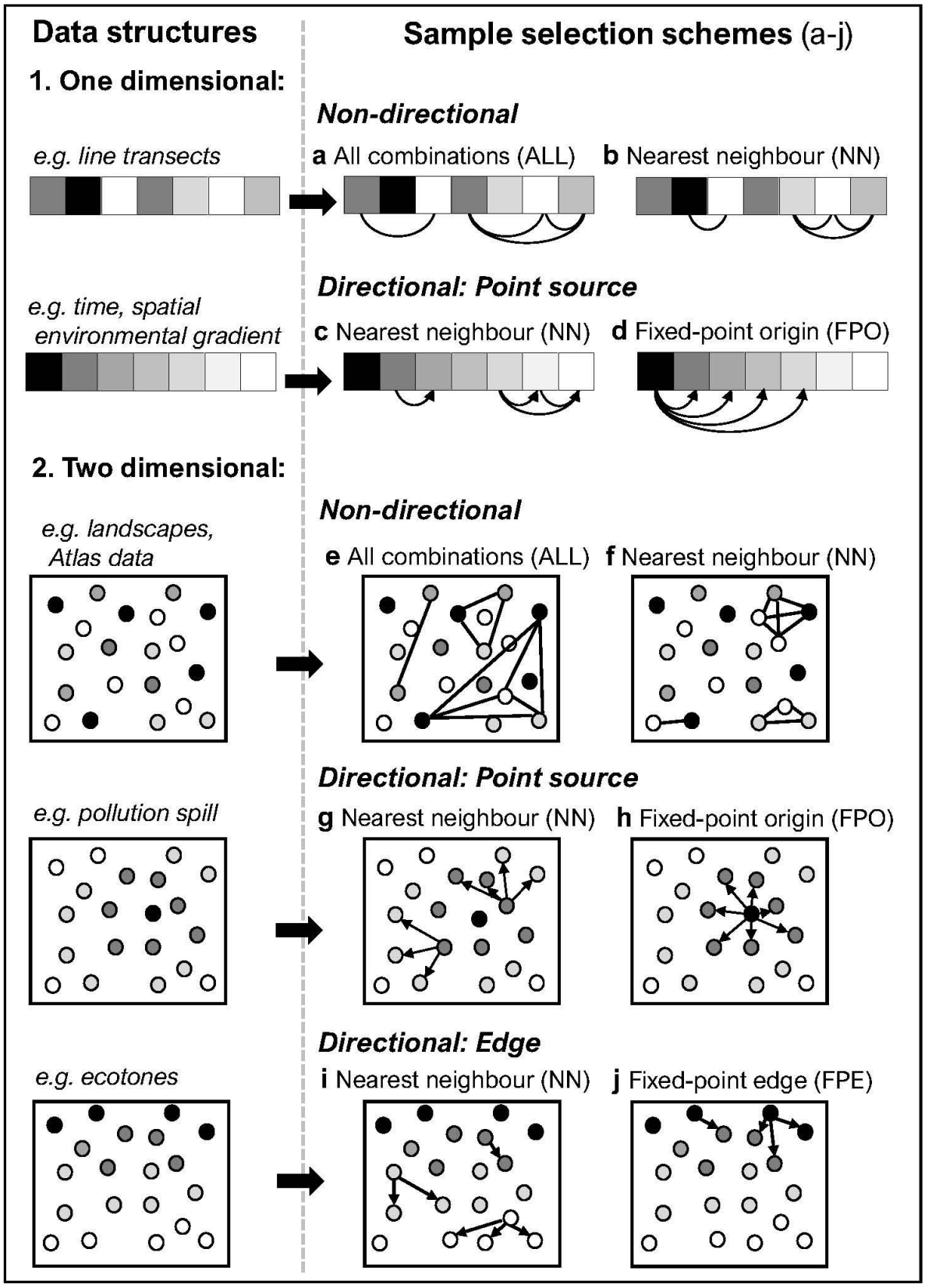
FIG. 4. Normalised Sørensen-equivalent zeta diversity decline for four data sets showing (a) how the number of shared species decreases with the zeta order. (b) The species retention rate using the zeta ratio, which shows the degree to which common species are more likely to be retained in additional sites or samples than rare ones with an increase in zeta order. (c, d) The form of decline against exponential (comparatively equal probability of species across sites) or power law fits (comparatively unequal probabilities of the occurrence of species across sites) (shown on log axes using normalised zeta). The best fit form for each dataset is shown with an asterisk (no best fit for Crop pests). Data sets: Crop pests (circles), DNAm disorder (squares), bioregion birds (triangles), soil metagenome (diamonds) (Table 2).

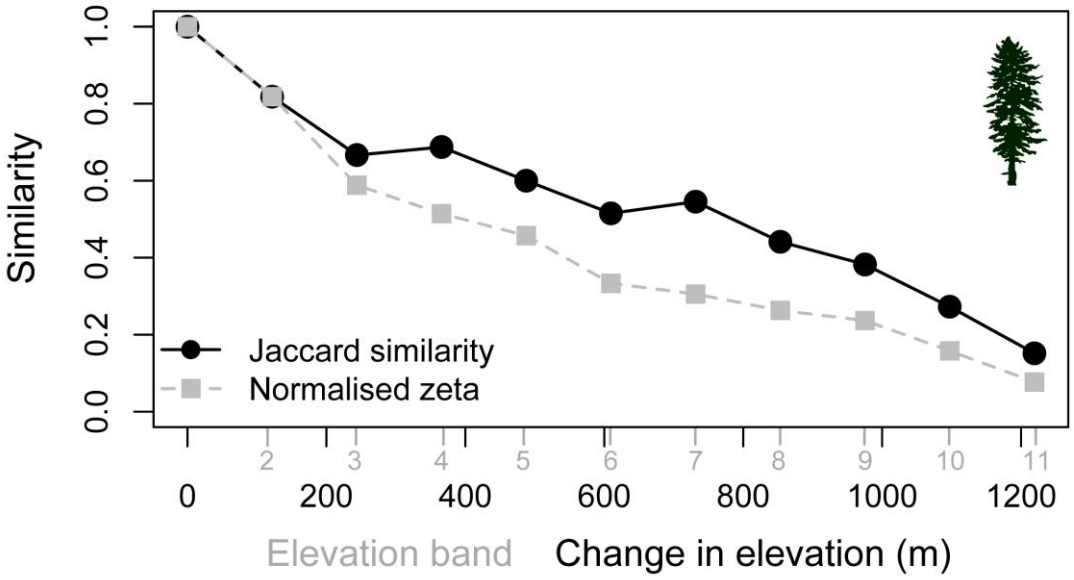
FIG. 5. Comparisons of zeta diversity decline between groups in three data sets (dashed lines = standard deviation), along with the retention rate using the zeta ratio, and exponential (E) versus power law (PL) fit (on log axes): (a) *Acacia* herbivores (beetles and bugs) across a

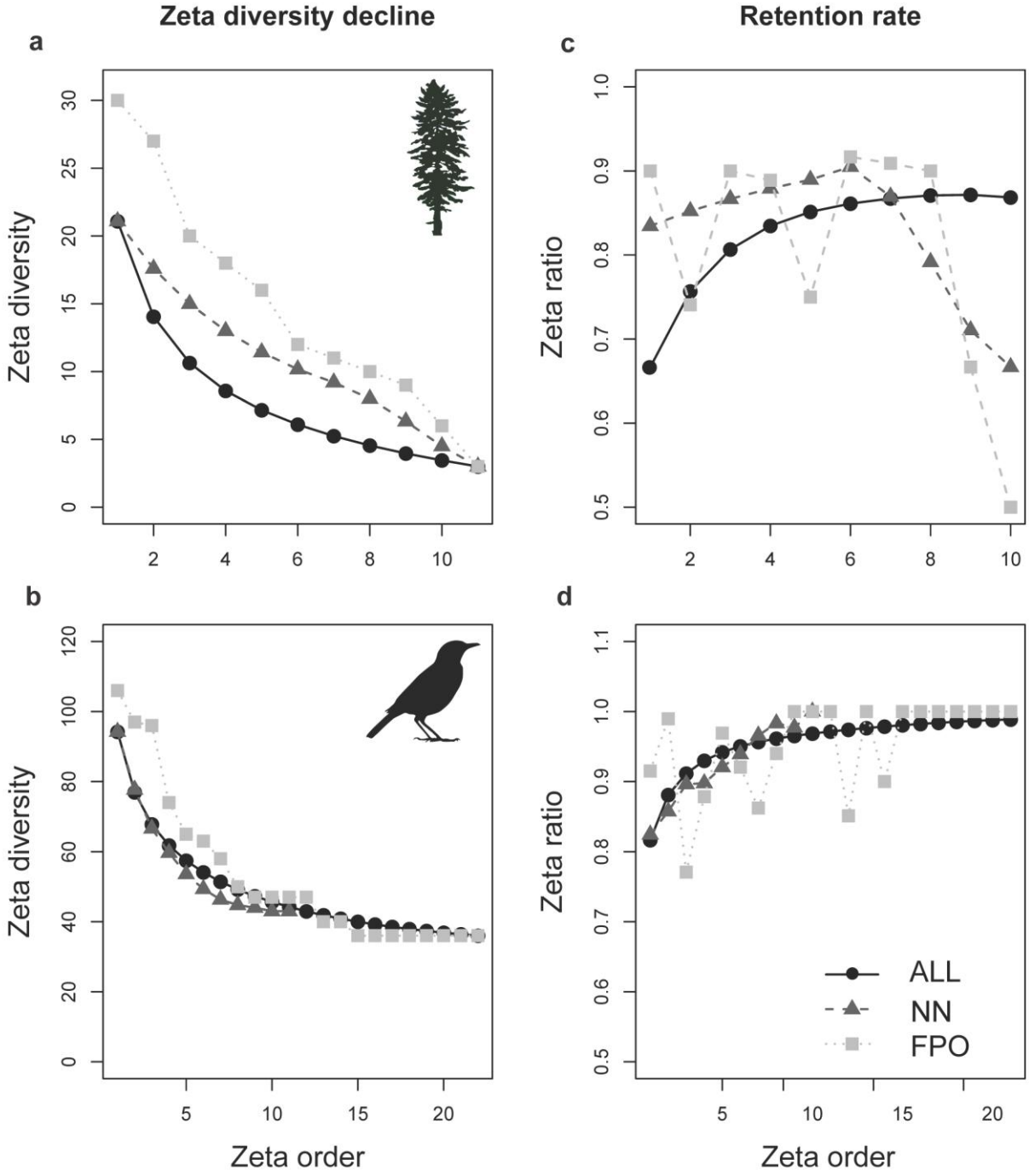
latitudinal range, (b) turnover in gall wasp interactions involving parasitoids or gall inquilines, and (c) DNA hypermethylation sites in patients with or without a disorder. The data sample selection scheme in all cases is 'ALL' combinations (Fig. 1). Best fit forms for datasets are shown with asterisks.

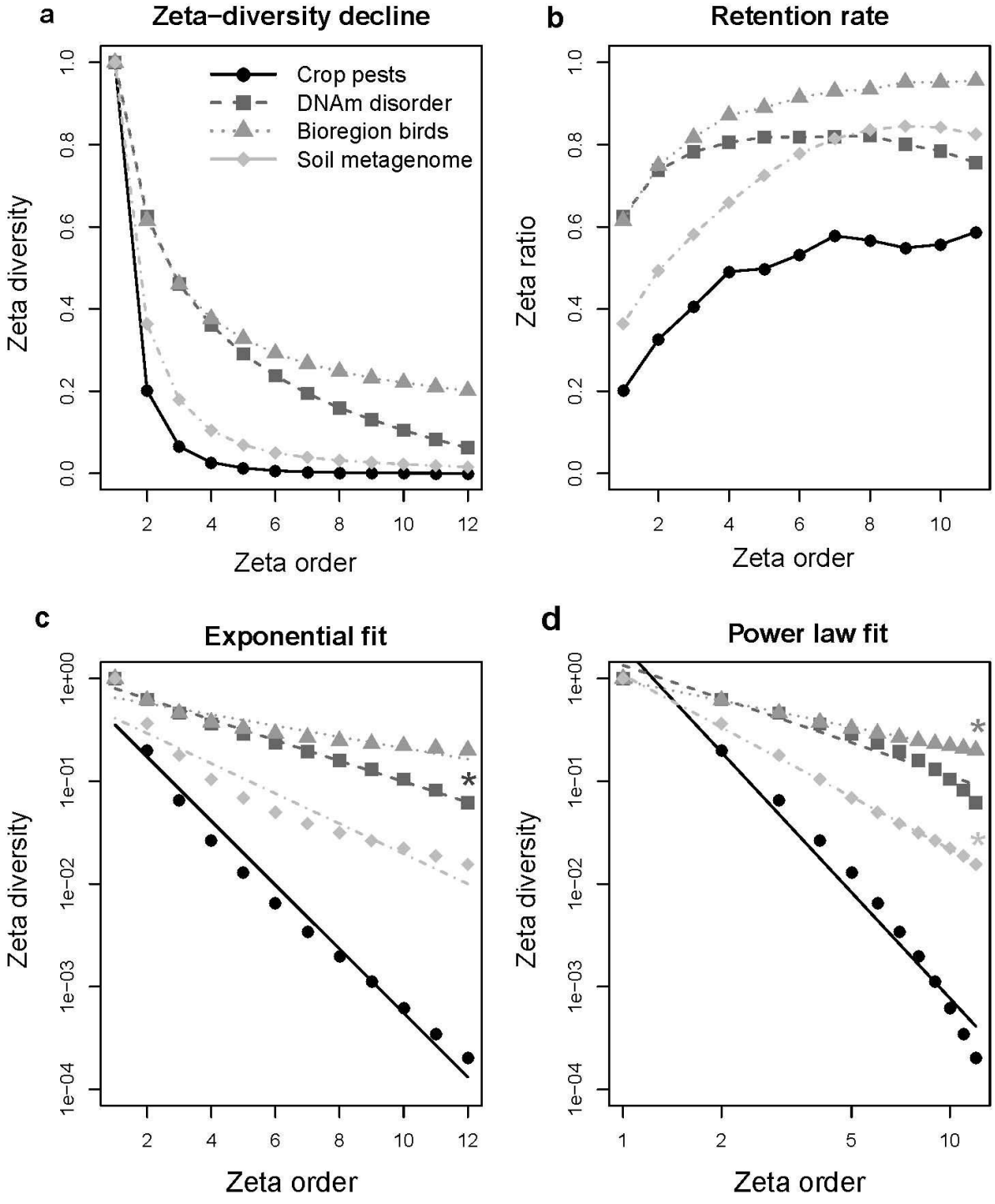
FIG. 6. Zeta diversity decay over space and time, for zeta orders 2 to 5, showing change in number of species shared with increasing distance between sites, or time between surveys (left). For completeness, the zeta diversity decline and the zeta ratio-based retention rates are shown on the right for the same data. (a-d) Alien and native plant species on Banks Peninsula (New Zealand); (e, h) microbial communities associated with kelp in two Australian marine biogeographic regions (NSW, New South Wales (east) and WA, Western Australia (west)); (i, l) temporal decay in bird communities in two catchments (Catchment 1 - below average rainfall; Catchment 2- average rainfall) over the course of a regional drought (1998-2003). Using the ALL combinations scheme (except in temporal birds, which uses DNN) (Fig. 1). Note that using mean distance for higher orders ($i > 2$) of zeta results in the increasingly narrow decay curve with increasing distance or time (see text). Note also that the ends of zeta decay curves, in particular the longer distance end, are usually associated with greater uncertainty because there are comparatively fewer sites this maximum distance apart than there are combinations of sites shorter distances apart (the same problem of unequal power across classes occurs in estimates of autocorrelation series, Legendre 1993). For orders $i > 2$, the distances between pairs of n sites are combined using, for example, mean distance; other options are the maximum extent of the area encompassed, or extent of occurrence, of the sites under consideration, or the maximum distance of sites apart. For example, in the kelp microbes (Fig. 6e, f) although compositional change in higher orders of zeta tended to mimic decay in ζ_2, they do so over a more narrow range of distances. This is a consequence of

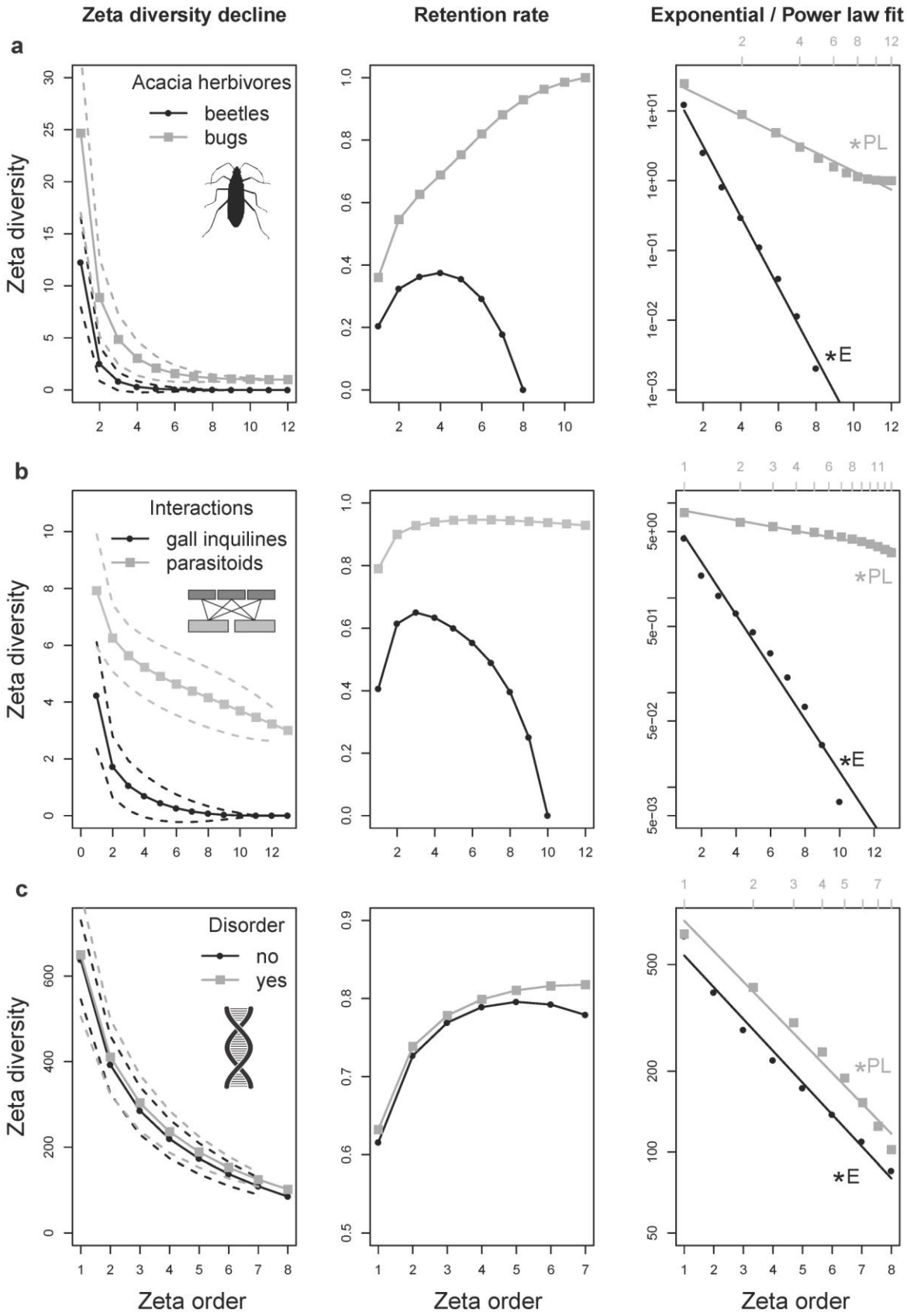
plotting decay against the mean distance across the *i* non-overlapping samples, with zeta of order *n* a single value for the average distance between all sites, and must be considered when interpreting the effect of distance on zeta diversity as the order increases (Latombe et al. 2017).

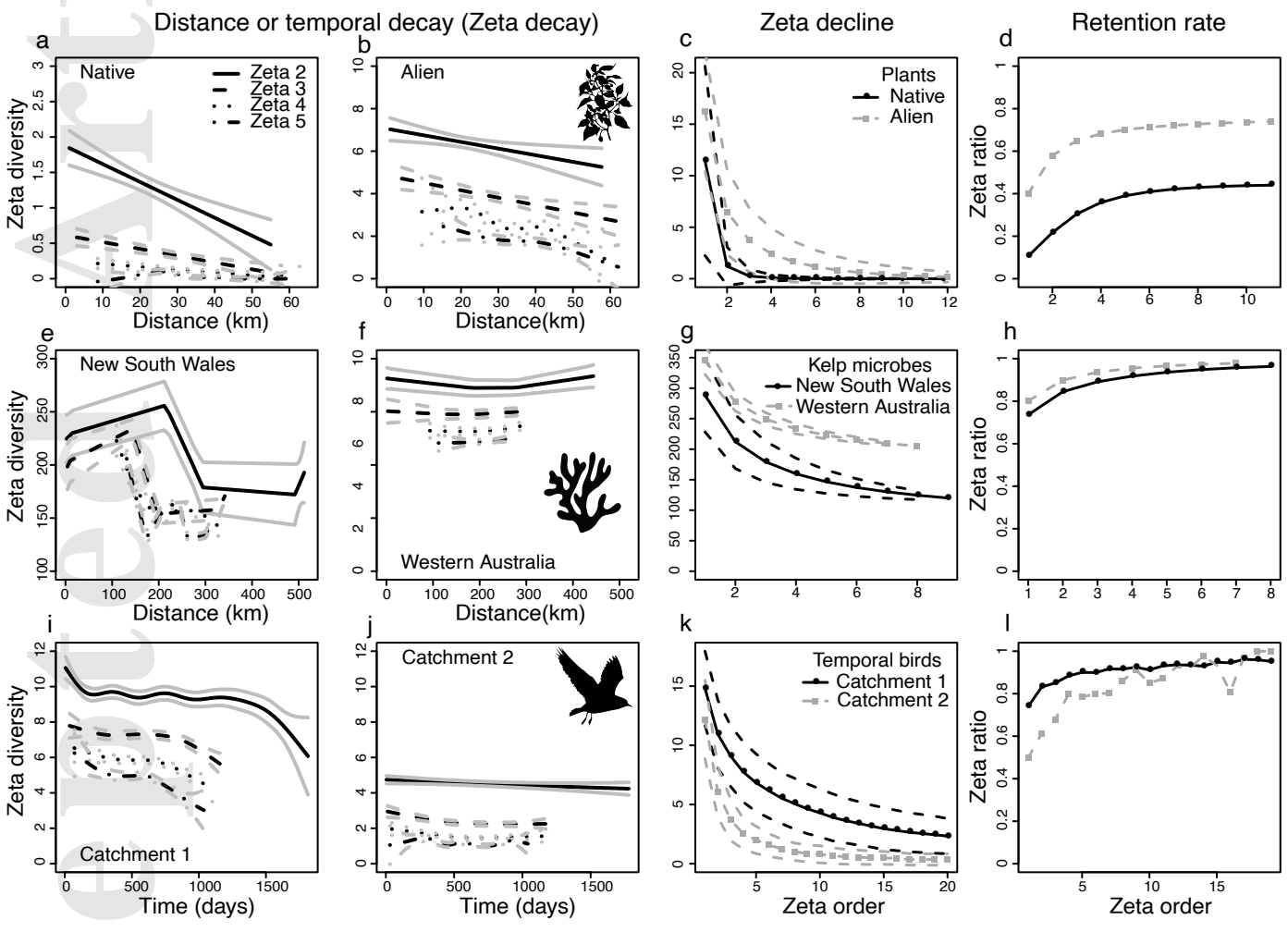











Distance or temporal decay (Zeta decay)
Zeta decline
Retention rate
a
Native
Zeta 2
Zeta 3
Zeta 4
Zeta 5
Zeta diversity
Distance (km)
b
Alien
Distance(km)
c
Plants
Native
Alien
d
Zeta ratio
e
New South Wales
Zeta diversity
Distance (km)
f
Western Australia
Distance(km)
g
Kelp microbes
New South Wales
Western Australia
h
Zeta ratio
i
Zeta diversity
Catchment 1
Time (days)
j
Catchment 2
Time (days)
k
Temporal birds
Catchment 1
Catchment 2
Zeta order
l
Zeta ratio
Zeta order